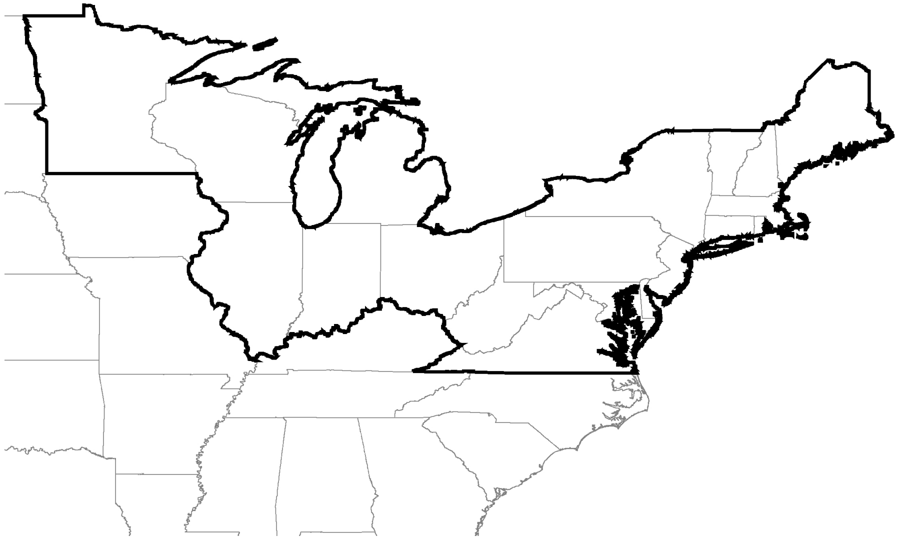
A hierarchical spatial count model with application to American Woodcock

Wayne Thogmartin, USGS Upper Midwest Environmental Sciences Center

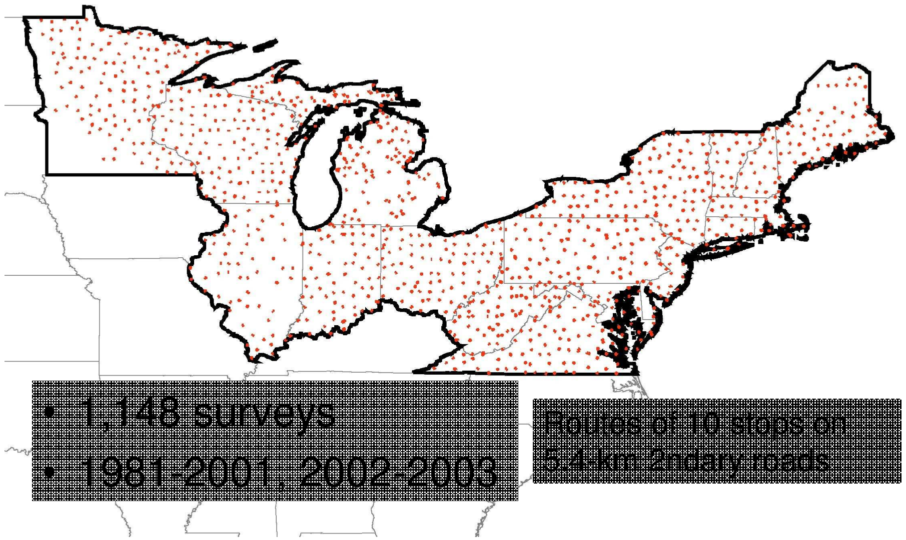
American Woodcock primary breeding range in the United States



Objective

- Our objective is to model and map predicted woodcock relative abundance across their primary breeding range in the United States.

American Woodcock Singing Ground Surveys



Survey Design

- Woodcock “peenting” surveys are annually conducted on secondary roads in the upper midwestern and northeastern United States.

Summary Statistics - Woodcock

- Mean count for 9,142 surveys (space x time) was 3.39 birds per survey (SD = 4.00)
- Zero counts comprised 27% of the surveys
- Median count was 2, and maximum count was 47
- 1,581 observers

Spatial Poisson Count Model

$$Z(s_i) = \mu(s_i) + \Sigma c_{ik}[Z(s_k) - \mu(s_k)] + \omega(s_i) + \gamma(s_i) + \varepsilon(s_i)$$

ω Observer effects: *observers count birds differently*

γ Year effects: *to accommodate observed decline in abundance*

μ Environmental effects

ε Extra-poisson variation

Spatial CAR (Conditional AutoRegression)

Spatial Poisson Count Model

- The expectation is treated as Poisson.
- Because observers count birds differently (e.g., older birders have a hard time hearing some species, novice birders have a hard time recognizing birds with unusual calls), we wish to adjust the counts to offset the effect of observer.
- We are using a time series of counts from a number of surveys. We can leverage this time series to inform our association of counts with habitat IF we control for annual variability and any sort of trends that may occur in the data (many birds are declining in abundance, and so it would be ‘unfair’ to compare counts from 1981 with those from 2001 if the species is in the midst of a decline).
- Environmental factors are included as a linear combination of variables derived from classified satellite imagery. These environmental factors will form the primary basis for mapping the predicted species abundance.
- Typically, the variance of counts exceeds the mean of those counts, so we have a term to soak up that extra-Poisson variation. This is generally not a serious issue as much of the extra-Poisson variation is ‘structural’ in nature, i.e., because of observers, routes, or years consistently leading to lower or higher counts than may be expected. This is adjusted for through hierarchical modeling (described shortly).
- We expect counts to be correlated over space, and so we model this correlation with a spatial 1st-order conditional autoregression.

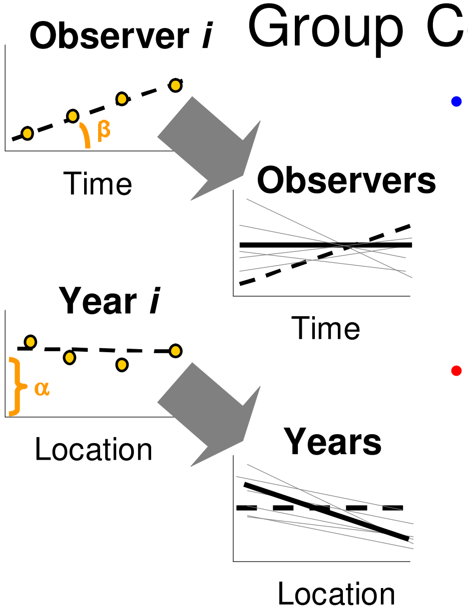
Hierarchical Modeling

- Bayesian: Data and prior specification used to identify a posterior distribution for parameter estimates (β)
  - Standardized Likelihood x Data = Posterior Probability
- Hierarchical: clustering of β for observer, year, and route effects because of group correlation
- Correlation may occur because of design, over time, and/or across **space**

Model Fitting

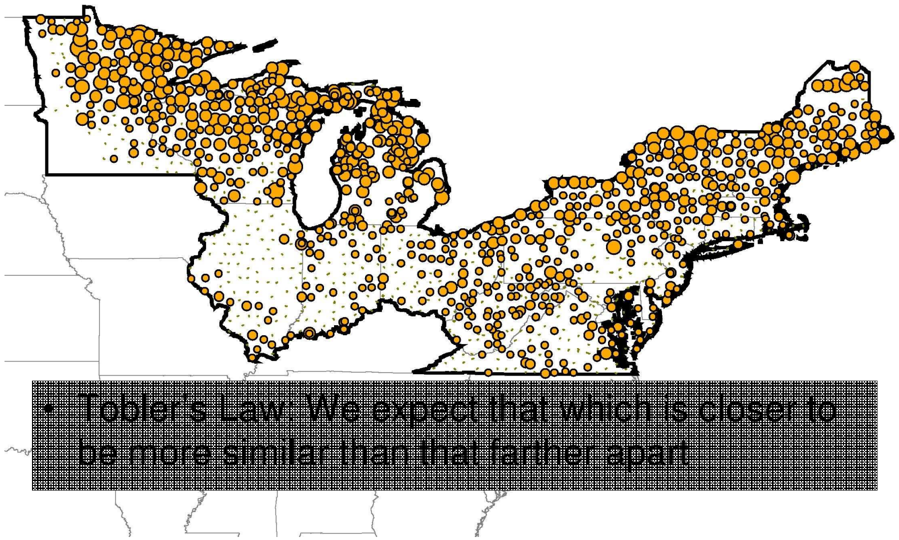
- To fit this model, the approach we employ is described as a hierarchical model. In this workshop, we're most interested in effects over space, but to identify those spatial effects free of the clutter of the sample design and the temporal correlation between survey's, we need to accommodate nuisance effects that would otherwise obfuscate the spatial effects.

Group Correlation



- Observer variability: Some observers will count woodcock similarly
- Temporal variability: Some years may be above the grand mean, others below

Mean Counts: Spatial Considerations

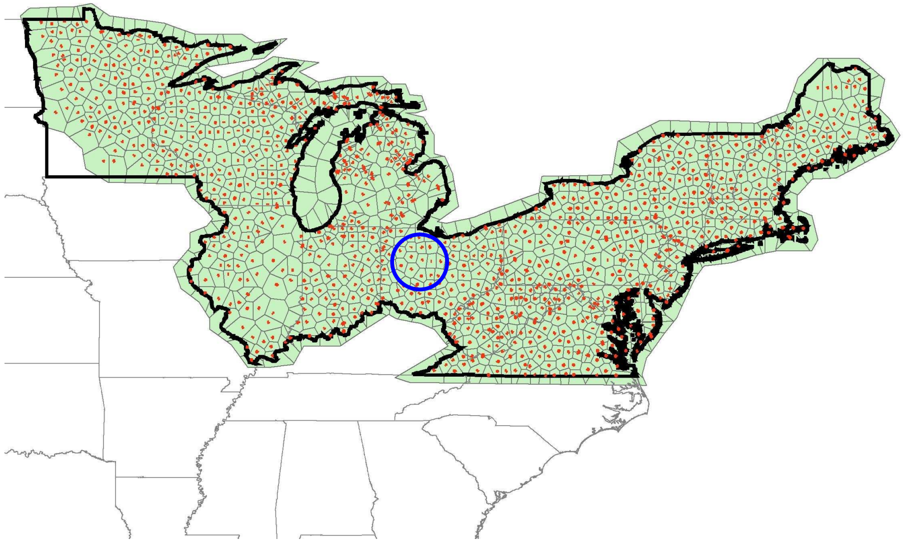


- Tobler's Law: We expect that which is closer to be more similar than that farther apart

Observed Counts

- An average of the point-specific time series shows that there is a general north-south gradient in woodcock abundance. Woodcock are more abundant in the north and less abundant in the south.
- This gradient results in sites near to one another being more similar than those farther from one another.
- We may wish to accommodate this spatial correlation to reduce the bias imposed on estimation of the slopes associated with the environmental factors. This correlation ostensibly describes environmental factors for which we have insufficient ability to map (i.e., understory plant composition, earthworm abundance, etc.).

Spatial Correlation: Lattice-based Solution



Irregular Lattice

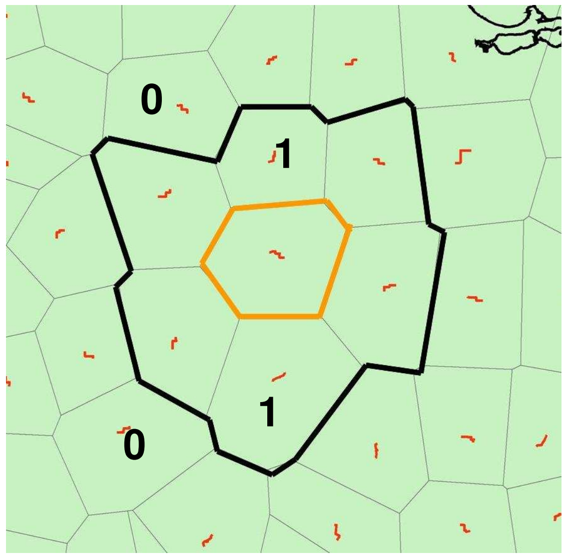
- We identify the domain of interest or influence around each route by tesselating the routes, creating an irregular lattice. This irregular lattice will be used to identify the neighborhood structure.

Neighborhood

- 1st order Conditional Autoregression

Value of *i* is akin to a weighted 'average' of surrounding cells

Surrounding cells weighted 1, distant cells weighted 0



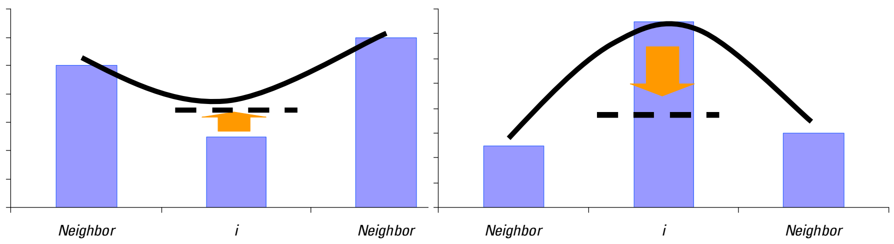
Shrinkage

- Observed counts will be more variable than the mean expectation. Shrinkage or smoothing gives a stable estimate of the pattern of the underlying expected counts, whereas the raw counts lead to a noisy or blurred picture of the true, unobserved count process.

Conditional Autoregression

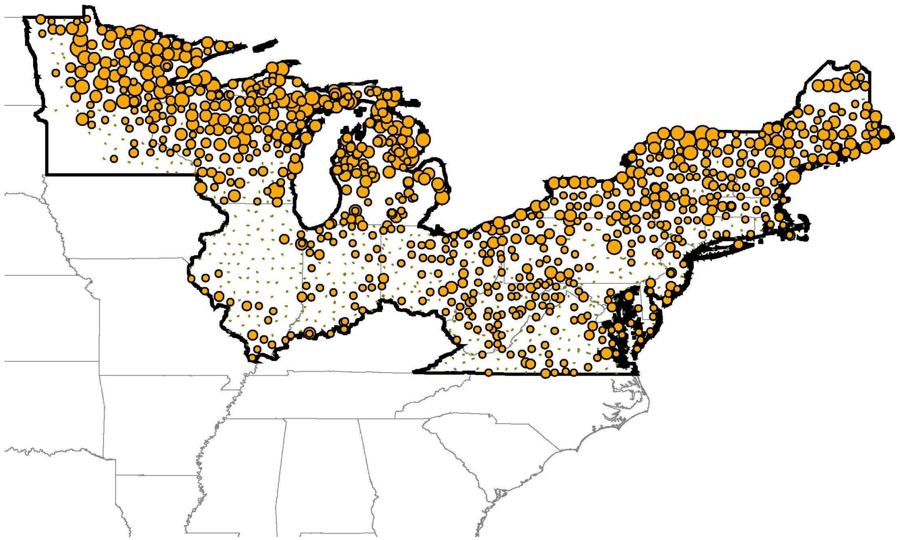
- Probability of observing a particular value at a given site is a conditional probability, i.e., it depends upon the values in the surrounding neighborhood
- Advantages:
  - Conservative
  - High Specificity (correctly classifying occurrences) even in sparse data situations

Smoothing

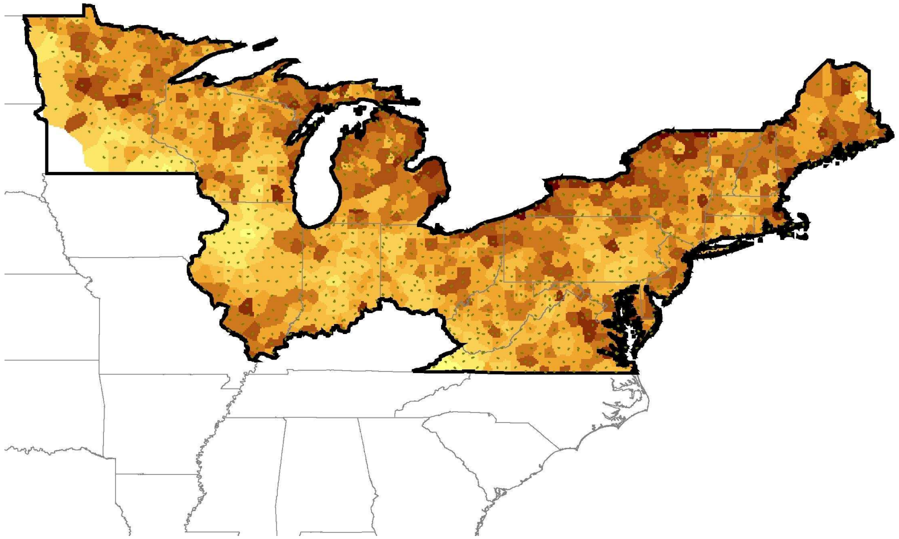


Shrinkage provides a stable estimate of the pattern of the underlying expected counts

Mean Counts



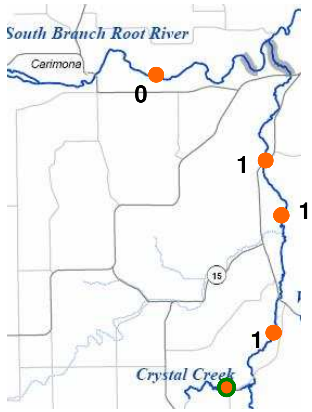
Smoothed Expectation



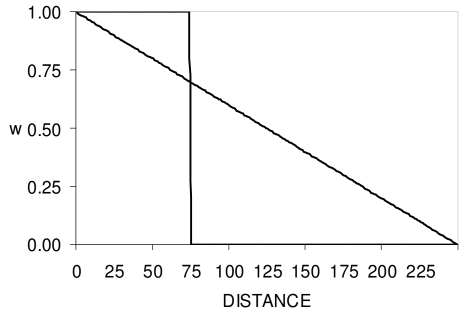
Alternatives to 1° CAR

- Ad hoc weighting

ω = 1 (1° & 2 °), 0 (>2 °)



Euclidean Distance-based weights



Weights determined by degree of interaction or similarity

Alternatives to 1° CAR

- We may wish to include not just our nearest neighbors, but also those neighbors in the surrounding ring immediately beyond the nearest neighbors; this would be a 2nd-order CAR.
- We also might want to use proximity as defined by a metric other than Euclidean distance. For instance, maybe only those points along a stream or road are considered part of the neighborhood and given a weight of 1, whereas all others are given a weight of 0.
- Others have used distance-based weightings, after having done semivariogram analyses to identify the degree of spatial correlation. These distance weightings can be 1 for all points within a certain distance (the range in geostatistical parlance) and 0 otherwise, or the 0-1 gradient can be continuous and reflect the distance from the point in some linear fashion.
- Regardless, symmetry needs to be observed. That is, if you are my neighbor, I am your neighbor.

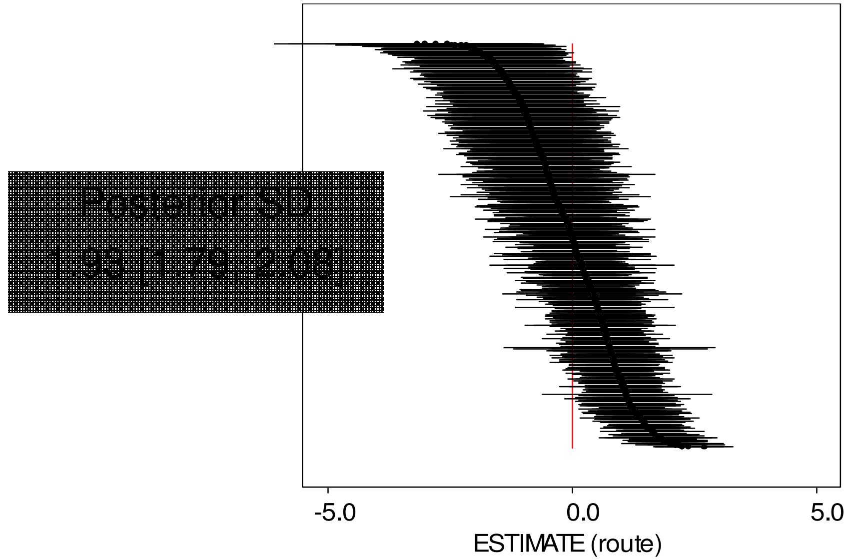
Parameter Estimates for μ for Models at 3 Spatial Scales

Variable	Finest Scale (350 ha)	Medium Scale (4,000 ha)	Coarsest Scale (106,000 ha)
INTERCEPT	0.02 (0.10)	0.07 (0.11)	0.06 (0.15)
START OF SEASON	-0.37 (0.17)	-0.33 (0.12)	-0.33 (0.16)
AGGREGATION INDEX	-0.29 (0.04)	-0.36 (0.05)	-0.26 (0.07)
HUMAN (%)	-0.22 (0.04)	-0.26 (0.04)	-0.15 (0.05)
GRASS (%)	-0.01 (0.05)	-0.21 (0.05)	-0.14 (0.07)
ASPEN (%)	0.09 (0.04)	0.12 (0.05)	0.20 (0.08)
TOPO CONVERGENCE	0.10 (0.04)	0.00 (0.05)	NA
SHRUB (%)	0.17 (0.11)	0.17 (0.11)	0.12 (0.14)
FOREST (%)	0.18 (0.05)	0.15 (0.05)	0.09 (0.05)
FOREST×FOREST(%)	NA	-0.01 (0.05)	NA

Environmental Factors

- We modeled woodcock at three spatial scales, and used an information-theoretic approach to averaging models within scale. We found little variability in woodcock response to the environment across scales. Woodcock were generally negatively related to the day of the year in which the growing season began, which may reflect the importance of earthworms to the woodcock diet. Woodock were also negatively related to landscapes in which forest, shrub, and field were aggregated into clumps as opposed to fine distributions among each other. The relation of forest and aspen were positive, but the importance of forest declined with the coarsening of scale, whereas the importance of aspen increased as the scale coarsened.

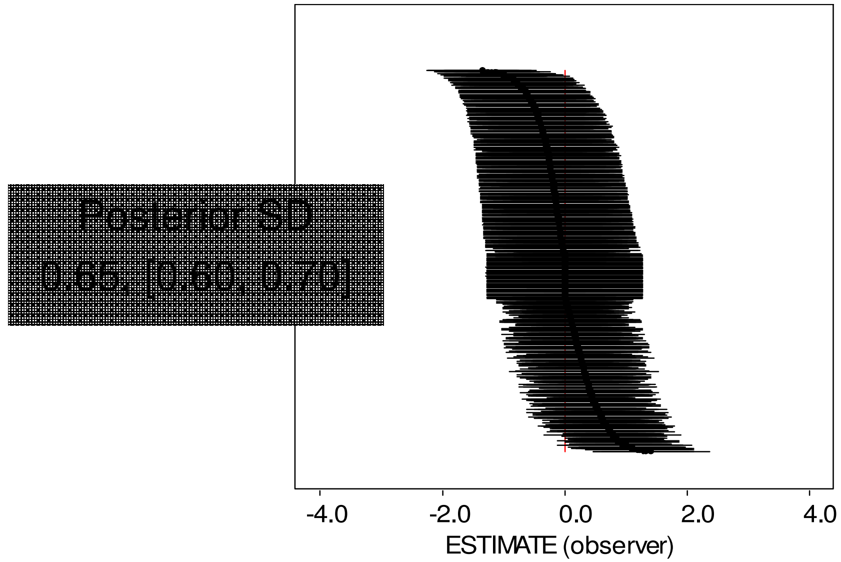
Route Random Effect



Route Random Effect

- A caterpillar plot of the individual route effects, ordered by route estimate, indicates a small number of routes reduce the expected counts relative to the predictions of the environmental variables, whereas a number of routes increase the expected counts relative to the predictions. These route-level reductions and increases are variability that we can not explain with the environmental variables we have identified in the course of our model.

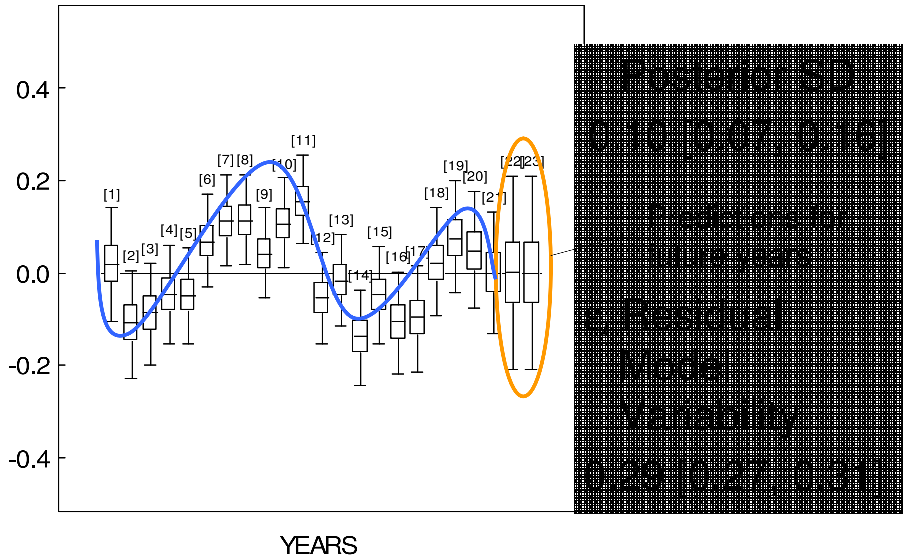
Observer Random Effect



Observer Random Effect

- Aside from a small number of observers who under- or over-counted relative to the other observers, most observers had little effect on the overall count expectation, indicating that we should have little concern in general for the effect of observers on surveys of woodcock.

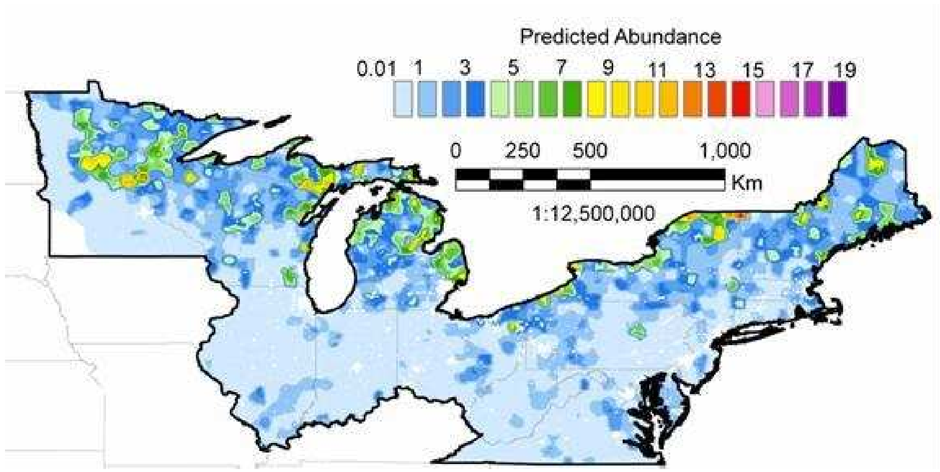
Year Random Effect



Year Random Effect

- May want to address the potential cyclicity with an AR(1) (i.e., an autoregressive term of lag 1); this may reduce the error variance around the out-years (2002 and 2003).

Predicted Woodcock Relative Abundance

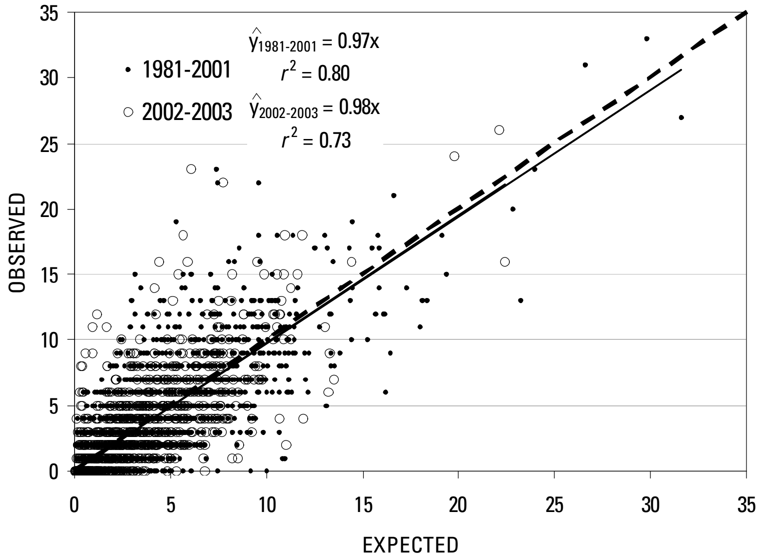


circa 1991

Mapped Predictions

- The result of mapping the environmental variables and the route effect together yields a map of predicted woodcock relative abundance.
- Because we treat the counts as Poisson, we must first exponentiate the linear combination of variables and slope estimates to map the count expectation. $e^{(\beta 1*X1 + ... + \beta k*Xk + \text{route effect } Z)}$

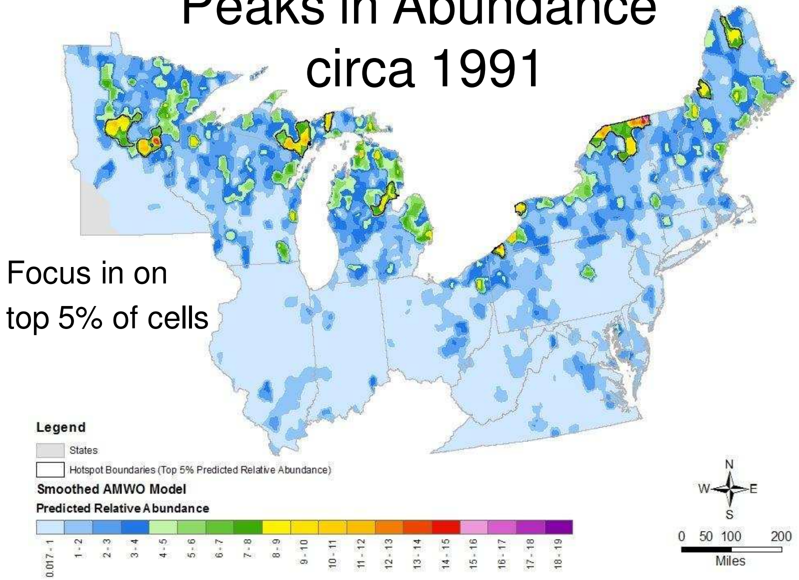
Model Evaluation



Model Evaluation

- Evaluation of the model by imputing values as determined by the final model structure (i.e., based upon the estimated model parameters [slopes]) indicated near one-to-one correspondence between the model predictions and the observed data for both those data withheld from model construction and data for the two years subsequent to the modeling effort.

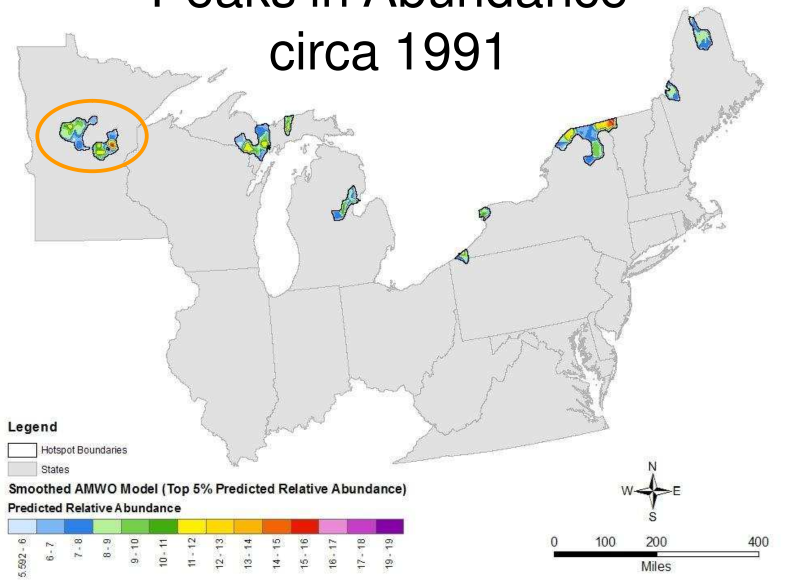
Predicted Woodcock Peaks in Abundance circa 1991



Management Application

- To increase the efficiency of conservation delivery, it would be best to manage the species where our efforts would do the most good for the most individuals of the population. Unless we are led to believe otherwise, that efficiency comes by conserving the species where it is most abundant. Thus, we use our map of predicted abundance to focus on specific areas of high or peak abundance.

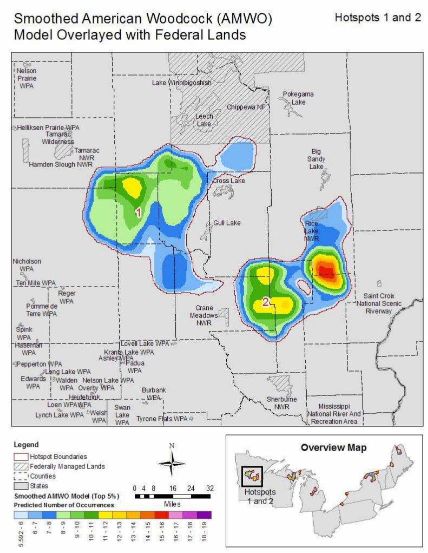
Predicted Woodcock Peaks in Abundance circa 1991

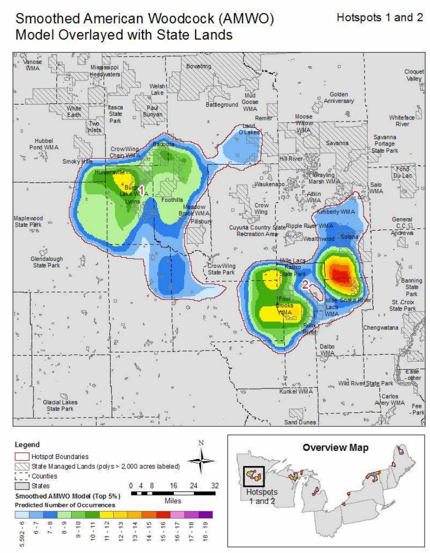


Management Application

- There are 10 such areas. These areas are the top 5% of the distribution in the expected counts.

Regional Conservation Planning





Management Application

- We found, from our analysis of the mapped predictions relative to the state and federal land management agencies (i.e., “the conservation estate”), that the proportion of the population occurring on private lands varied between 70.5% in Minnesota to 94.1% in Maine, with a grand mean of 79.9%. The proportion of the predicted population was 7.2% on federal land and 12.9% on state land, which was marginally higher than the proportion of the area under federal and state management (6.4% and 11.4%, respectively).

- We plotted our predictions against data layers describing the land management context with the idea that land managers and private lands biologists can effectively direct species-specific conservation efforts to those specific areas where the species is high in abundance. We can also use these sorts of maps to direct research activities, to better learn why species in these areas are highly abundant. We may also be able to use constituent aspects of the model to identify areas where the species can be most effectively increased by simple modification of the landscape (i.e., if we affect certain management practices in areas where the species occurs, might we see better bang for our buck in some areas rather than other areas; are there limiting factors that we can not overcome regardless of our management efforts [e.g., climate (start of the growing season) can not be managed, but only accommodated]).

Questions?

- For more information: http://www.umesc.usgs.gov/terrestrial/migratory_birds/bird_conservation/amwo_american_woodcock.html

- wthogmartin@usgs.gov